5-Ч

УДК 004.82:004.62:314.1

О.Р. Чертов
Національний технічний університет України «Київський політехнічний інститут», м. Київ
chertov@i.ua

Моделювання метаданих опису статистичної інформації

У статті розглядаються питання, які виникають під час моделювання метаданих, що використовуються для опису різноманітної статистичної інформації. Конкретизується відома модель статистичних бізнес-процесів та пропонується нова семіотична модель метаданих. В них, зокрема, враховані аспекти забезпечення знеособленості та конфіденційності статистичної інформації, а також її захист від розкриття при застосуванні методів інтелектуального аналізу даних.

Вступ

Проект розвитку системи державної статистики України для моніторингу соціально-економічних перетворень за підтримки Міжнародного банку реконструкції та розвитку [1] направлений на забезпечення єдиної технології збору, обробки, накопичення, представлення і розповсюдження статистичної інформації, гарантування надання надійних, своєчасних та точних даних і їх ефективне використання для прийняття рішень. Роз’язання завдань вказаного проекту можливе лише за умови створення Інтегрованої системи обробки статистичних даних (ІСОСД), впровадження котрої заплановано на 2012 р.

Сучасний підхід до створення інформаційних систем в галузі статистики був закладений близько 10 років тому [2]. З того часу щороку під егідою Статистичного Управління Європейських Співтовариств Євростат (Eurostat) проводяться спеціалізовані семінари, що носять з 2004 р. назву «Meeting on management of statistical information systems», на котрих обговорюються різноманітні аспекти розробки та впровадження відповідних систем. Аналіз робочих документів цих семінарів показує, що побудова новітніх інформаційних систем в галузі статистики істотно спирається на суцільне впровадження концепції підтримки метаданих.

Процес розробки загальної структури системи статистичних метаданих було ініційовано в 2004 р. під час спільного робочого засідання Європейської економічної комісії ООН (UNECE), Євростату і Організації економічної співпраці та розвитку (OECD), присвяченого питанням інформаційної системи статистичних метаданих (METIS). Наразі вироблені стандарти і рекомендації стосовно статистичних метаданих, які вже застосовуються на практиці національними і міжнародними статистичними організаціями. До них, зокрема, відноситься комплекс технічних стандартів та настанов SDMX (Statistical Data and Metadata eXchange) [3], настанова щодо звітування та представлення даних та метаданих (The OECD Data and Metadata Reporting and Presentation Handbook) [4], стандарт загальної моделі статистичних бізнес-процесів (Generic Statistical Business Process Model) [5] та рекомендації щодо роботи із статистичними метаданими в мережі Інтернет (UNECE Guidelines for Statistical Metadata on the Internet) [6].

Але всі зазначені нормативні чи рекомендаційні документи розраховані або на опис архітектури та загальних статистичних бізнес-процесів [3], [5], або на опис метаданих з точки зору взаємодії з іншими (нестатистичними) інформаційними системами [4], [6]. В той же час для практичного впровадження інтегрованої статистичної системи, керо-

ваної метаданими, потрібна також детальна проробка моделі метаданих, що буде використовуватися для опису статистичної інформації та конкретного процесу її обробки.

Також необхідно явно виділити питання забезпечення конфіденційності інформації, бо вони мають винятково важливе значення в системах надання чи розповсюдження статистичної інформації. Страсбурзька Конвенція «Про захист фізичних осіб в частині автоматичної обробки персональних даних» [7] передбачає особливий статус статистики і у зв'язку з цим проголошує необхідність обмеження права доступу до статистичної інформації та до комп'ютерних файлів і баз даних, що використовуються в статистичних чи науково-дослідних цілях. Зростання в останні десятиліття важливості забезпечення знеособленості та конфіденційності інформації викликано цілим рядом причин [8-9]:

– збільшенням попиту на дані невеликих територіально-адміністративних одиниць (для об'єктивного формування державної регіональної політики);

– поширенням вибіркових обстежень та розповсюдженням мікроданих;

– збільшенням кількості перехресних джерел інформації (статистичних, адміністративних тощо), що підвищує ризик розкриття приватної чи конфіденційної інформації шляхом зіставлення даних;

– розширенням можливостей інформаційних систем, зокрема, швидким розвитком методів інтелектуального аналізу даних (data mining), що дозволяють виявляти структури та закономірності в статистичних даних.

У зв'язку з цим **метою даної роботи** є розробка моделі метаданих опису статистичної інформації, що враховувала б особливості обробки даних в статистичній галузі та явним чином підтримувала забезпечення конфіденційності інформації, що зберігається.

Постановка завдання. Є статистична інформація, яка збирається, зберігається, оброблюється та розповсюджується державними органами статистики. Цю інформацію формально можна розглядати як набір структурованих записів $\{A_i\}$, що надходять чи оброблюються в певний момент часу. Потрібно розробити модель метаданих, тобто даних, які описують записи $\{A_i\}$, котра б дозволила, з одного боку, представити всі процеси обробки даних, що наразі використовуються органами державної статистики України, а з іншого боку, підтримувала б забезпечення конфіденційності даних із записів $\{A_i\}$, як відносно окремих респондентів, так і їх груп. Відповідний захист статистичних даних повинен також враховувати загрози розкриття інформації шляхом застосування методів інтелектуального аналізу даних.

Мікродані, макродані, метадані

З самого початку свого розвитку інформаційні системи в галузі статистики підтримували збереження та обробку даних двох видів: мікроданих та макроданих.

Під *мікроданими* розуміють дані про об'єкти статистичного спостереження. Наприклад, стать чи дату народження респондента, кількість респондентів в домогосподарстві тощо.

Макродані – це оцінні значення статистичних показників, що відносяться до групи (сукупності) об'єктів. Вони тим чи іншим чином агрегують інформацію про групу однотипних об'єктів. Наприклад, середня кількість членів в домогосподарствах певного району, кількість національностей, що включені до найбільш численних в певному регіоні тощо.

Але як для опису мікроданих, так і для опису макроданих можна використовувати метадані. Існує багато означень метаданих. Наприклад, у звіті [10] їх наведено рівно 27, включаючи найлаконічніше: метадані – це дані про дані.

Ми в подальшому будемо користуватися наступним означенням: *статистичні метадані* – це структуровані дані, що характеризують різні аспекти статистичних даних, наприклад, їх зміст, точність, доступність тощо.

Модель статистичних бізнес-процесів

Статистична спільнота вже тривалий час займається питаннями побудови загальної моделі статистичних бізнес-процесів (статистичного циклу). Мета цих робіт полягає у формуванні для всіх статистичних організацій узгодженої стандартної термінології та опису застосовуваних бізнес-процесів, необхідних для виробництва офіційної статистики. Це сприятиме як стандартизації процесів обробки та розповсюдження статистичної інформації, так і гармонізації статистичних обчислювальних інфраструктур, полегшенню процесу обміну компонентами відповідного програмного забезпечення.

Наразі можна виділити чотири основні міжнародні моделі та стандарти, що певним чином пов’язані з моделлю статистичних бізнес-процесів:

1) керівництва та рекомендації щодо структури інформаційних систем для національних та міжнародних статистичних установ [2] містять узагальнену модель етапів та процесів для систем обробки даних статистичних спостережень;

2) модель Євростату «життєвий цикл даних» («Cycle de Vie des Donées») [11] надає повний перелік концепцій, структур метаданих та інформаційних технологій для застосування в усіх галузях статистики;

3) комбінована модель життєвого циклу, розроблена в рамках Ініціативи з документування даних (ІДД) [12] – відкритого міжнародного проекту зі створення стандарту для технічної документації, яка описує соціологічні дані;

4) загальна модель статистичних бізнес-процесів (ЗМСБП) [5], детальніше описана нижче.

Хоча за формою представлення моделі відрізняються поміж собою, проте їх зміст – досить близький. Модель ЗМСБП з точки зору виконання статистичних задач здебільшого перекриває можливості трьох інших моделей. Наприклад, модель ІДД включає, на перший погляд, унікальний процес «Перепрофілювання», тобто вторинного використання набору даних, яке спочатку не передбачалося на етапах планування та збору даних. Однак в моделі ЗМСБП цей процес можна зіставити з двома підпроцесами: «перевірки доступності існуючих даних і їх використання в разі можливості» (на етапі «Уточнення потреб») та «інтеграції даних» (на етапі «Обробка»).

Загалом модель ЗМСБП включає в себе опис 9 етапів, кожен з яких складається з різної кількості підпроцесів (від трьох до восьми) – табл. 1 [5, с. 9]. Назви підпроцесів досить повно характеризують основні операції з обробки даних, які на них виконуються.

Таблиця 1 – Загальна модель статистичних бізнес-процесів

1	2	3	4	5	6	7	8	9
Уточнення потреб	Проектування	Формування	Збір	Обробка	Аналіз	Поширення	Архів	Оцінка
1.1 Визначення потреб у інформації	2.1 Проектування результатів	3.1 Створення інструменту збору даних	4.1 Вибірка	5.1 Інтеграція даних	6.1 Підготовка проекту вихідних даних	7.1 Оновлення вихідних систем	8.1 Визначення правил архівування	9.1 Збір вхідних даних для оцінки
1.2 Консультації та підтвердження потреб	2.2 Проектування описів змінних	3.2 Створення (збільшення) компонентів процесу	4.2 Початок збору	5.2 Класифікація і кодування	6.2 Підтвердження вихідних даних	7.2 Виробництво продуктів поширення	8.2 Ведення архівного сховища	9.2 Проведення оцінки
1.3 Визначення завдань вихідних даних	2.3 Проектування методів збору даних	3.3 Формування робочих процесів	4.3 Збір даних	5.3 Перегляд, підтвердження і редагування	6.3 Дослідження та пояснення	7.3 Випуск продуктів поширення	8.3 Збереження даних і метаданих	9.3 Узгодження плану дій

Продовження табл. 1

1.4 Визначення понять	2.4 Проектування методів визначення меж та вибірки	3.4 Випробування системи виробництва	4.4 Завершення збору	5.4 Використання розрахункових даних	6.4 Контроль над відкриттям інформації	7.4 Просування продуктів поширення	8.4 Знищення даних і відповідних метаданих	
1.5 Перевірка доступності даних	2.5 Проектування методів обробки даних	3.5 Випробування статистичного бізнес-процесу		5.5 Отримання нових змінних та статистичних одиниць	6.5 Оформлення вихідних даних	7.5 Здійснення підтримки користувачів		
1.6 Підготовка біз-нес-справи	2.6 Проектування робочого процесу	3.6 Завершення формування системи виробництва		5.6 Обчислення значень				
				5.7 Обчислення сукупностей				
				5.8 Оформлення файлів даних				

Враховуючи практику проведення статистичних спостережень в Україні, що склалася за роки незалежності, і явно вказавши процеси, які забезпечують знеособленість та збереження конфіденційності статистичної інформації, модель ЗМСБП можна конкретизувати, виділивши в ній 14 процесів та територіальні рівні статистичної обробки (рис. 1). На місцевому рівні маємо 498 районних та міських відділів статистики. Офіційними органами статистики України на регіональному рівні є Головне управління статистики в Автономній Республіці Крим, 25 головних управлінь статистики в областях і місті Києві, а також Управління статистики в місті Севастополі. Центральний рівень забезпечує Державний комітет статистики України, куди ми також віднесемо Головне міжрегіональне управління статистики в місті Києві (реорганізований в 2000 р. колишній Головний обчислювальний центр).

Зазначимо, що в запропоновану модель статистичних бізнес-процесів не включені загальносистемні процеси на кшталт стратегічного планування, управління людськими ресурсами, управління фінансами, управління організаційною структурою тощо. Також у зв'язку з тим, що питання уточнення потреб, проектування та формування статистичного процесу мають, насамперед, методологічний характер та здійснюються, як правило, один раз для конкретного процесу, то вони були винесені за межі моделі, що розглядається. З іншого боку, систематичні процеси обробки статистичних даних були ґрунтовно деталізовані в порівнянні з моделлю ЗМСБП.

З точки зору наближення статистичних даних до їх користувача дуже важливим є впровадження доступу до пошуку необхідної інформації:

– за допомогою нерегламентованих запитів до інформаційно-облікової статистичної системи;

– через аналітичні запити до інформаційно-аналітичної статистичної системи;

– шляхом застосування інтелектуального аналізу даних для автоматизованого пошуку в них певних закономірностей.

Тому в запропонованій на рис. 1 моделі був явно виділений процес «Пошуку», який в моделі ЗМСБП розосереджено по процесах «Обробка», «Аналіз» та «Поширення».

Забезпечення конфіденційності статистичної інформації визначається розділом V Закону України «Про державну статистику» [13]. Зокрема, в ньому зазначається, що забороняється «поширення статистичної інформації, на підставі якої можна визначити конфіденційну статистичну інформацію щодо конкретного респондента».



Рисунок 1 – Процеси та територіальні рівні статистичної обробки в Україні

Замість підпроцесу «6.4. Контроль над відкриттям інформації» моделі ЗМСБП в пропонованій моделі введено два загальні процеси: «Знеособлення» та «Забезпечення конфіденційності».

Під знеособленням персональних даних ми розуміємо вилучення відомостей, які дають змогу ідентифікувати особу (наприклад, номер та серія паспорту, прізвище, ім'я та по батькові тощо). Термін «знеособлення даних» було введено в модельному законі «Про персональні дані», що був прийнятий 16 жовтня 1999 р. на 14 пленарному засіданні Міжпарламентської асамблеї держав-учасниць СНД, принципи і норми котрого тим чи іншим чином мають бути включені до національних законодавств зазначених держав.

В забезпеченні конфіденційності статистичної інформації ми виділяємо три підпроцеси:

1) визначення мінімально допустимої кількості респондентів, інформація про котрих може бути надана (у вихідній таблиці, у відповіді на аналітичний запит тощо), наприклад, під час обробки даних Всеукраїнського перепису населення 2001 р. ця кількість дорівнювала 3000 осіб;

2) забезпечення індивідуальної анонімності даних [9] – окрім здійснення знеособленості, існують й інші методи, що дозволяють приховати інформацію про конкретного респондента, наприклад, зменшення рівня деталізації інформації про посаду особи – не «прокурор району», а «працівник правоохоронних органів»; дані методи застосовуються в основному під час підготовки мікрофайлів – наборів даних, що містять інформацію про вибіркову кількість респондентів;

3) забезпечення групової анонімності даних [14], бо при наданні користувачам доступу до первинних чи агрегованих даних у вигляді мікрофайлів з'являється можливість виявлення характерних ознак розподілу даних, наприклад, максимуми територіального розподілу концентрації військових можуть вказувати на наявність військових баз на відповідній території.

Розширена семіотична модель метаданих опису статистичної інформації

В роботі [15] була висунута ідея побудови моделі метаданих для опису статистичної інформації, яка виходить з визначення трьох проекцій тривимірного представлення метаданих (табл. 2). Беручи до уваги класичний поділ семіотики на синтаксис, семантику та прагматику, ми пропонуємо такі моделі називати *семіотичними*.

Таблиця 2 – Проекції (семіотичні рівні) опису статистичних метаданих

№	Семіотичний рівень	Направленість опису метаданих	Коротке пояснення
1	синтаксична проекція	на представлення	метадані є фізичним представленням метаінформації
2	семантична проекція	на зміст	метадані надають інформацію про дані
3	прагматична проекція	на цілі використання	метадані як дані, необхідні для підготовки та використання статистичних даних

Наведене в табл. 2 представлення метаданих є типовим для різноманітних семіотичних моделей. Однак воно потребує розширення, бо не враховує явним чином ще одного (четвертого) рівня – рівня конфіденційності статистичної інформації, що зберігається чи оброблюється. Інформаційні системи в галузі статистики завжди мали справу з діалектичною протилежністю точності та конфіденційності даних: саме первинні мікродані можуть забезпечити найбільшу точність статистичної інформації, але ж вони одночасно мають найнижчий рівень знеособленості.

В роботах [2] і [15] синтаксична та семантичні проекції були деталізовані для опису макроданих, які традиційно представляються в статистиці у вигляді таблиць. Автор цього підходу Бо Сунгрен (Bo Sundgren) виділяє 4 компоненти (виміри) опису:

– α: об’єктний (респондентський) вимір (the dimension of population and scope);
– β: підрахувальний вимір (the dimension of measurement and summation);
– γ: класифікаційний вимір (the dimension of classification);
– τ: часовий вимір (the dimension of time).

Приклади відповідних компонент наведені в табл. 3.

Таблиця 3 – Компоненти метаопису макроданих

	компоненти макроданих			
назва компоненти	α-компоненти	β-компоненти	γ-компоненти	τ-компоненти
синтаксична проекція	стовпчик чи рядок таблиці	чарунки таблиці	стовпчик чи рядок таблиці	стовпчик, рядок чи заголовок таблиці
семантична проекція (сенс компоненти)	Які об’єкти підраховуються?	Яка величина підраховується?	За якими властивостями (характеристиками) йде розподіл об’єктів?	Який часовий ряд береться, тобто з якого часового діапазону і з якою періодичністю?
приклад	резиденти України в певний момент часу	кількість/1000	за місцем проживання	щорічно, з 01.01.2000 по 01.01.2010

Всі статистичні метадані традиційно [15, с. 16-19] об’єднуються в групи. Але, на наш погляд, класифікація, наведена в [15], знову ж таки, не враховує важливість та особливість питань забезпечення конфіденційності статистичної інформації, а тому вимагає розширення, що знайшло відображення в додаткових рядках, виділених в табл. 4 курсивом.

Таблиця 4 – Групи статистичних метаданих

Призначення метаданих	Назва групи метаданих	Умовний код / назва підгрупи (якщо є)	Пояснення
для користувачів	загальні (глобальні)	К1	опис, тезаурус, індекси
	декларативні	К2.1 / сенс	для оцінки релевантності даних
		К2.2 / точність	стосовно розрахунку чи оцінки
		К2.3 / доступність	як отримати доступ до даних
	процесо-орієнтовані	К3	застосовані моделі, процедури вибірки, кодування, редагування, верифікації тощо
для постачальників даних	загальні (глобальні)	П1	застосована методологія, близькі за методикою статистичні задачі
	зворотний зв’язок	П2	чи задоволені користувачі отриманими даними і наскільки, якщо задоволені
	процесо-орієнтовані	П3	аналогічно К3
	орієнтовані на конфіденційність	*П4.1*	*знеособленість та індивідуальна анонімність*
		П4.2	*групова анонімність*
пов’язані з програмним забезпеченням	фактографічні	С1	назви таблиць, формати представлення даних тощо
	алгоритмічні	С2	алгоритми обчислення та їх опис

За теорією моделювання процесів [5, с. 3], кожен процес повинен мати певну кількість чітко виражених атрибутів, включаючи: вхідні дані, вихідні дані, мету (з додатковими характеристиками), власника, керівництва (наприклад, інструкції з використання та документація), дозвільні елементи (люди та системи), механізми зворотного зв'язку.

Зазначені атрибути в обов'язковому порядку повинні бути враховані в моделі опису статистичних даних, якщо вона має використовуватися на практиці.

Наведені вище міркування, включаючи конкретизовану модель статистичних бізнес-процесів, дозволяють запропонувати нову, розширену семіотичну модель метаданих для опису статистичної інформації.

Для ілюстрації застосування цієї моделі наведемо в табл. 5 фрагмент опису поняття «статистичне спостереження», як це пропонується робити в ІСОСД України.

Висновки

В роботі здійснена конкретизація загальної моделі статистичних бізнес-процесів [5] для статистичної галузі України та запропонована нова розширена семіотична модель метаданих для опису статистичної інформації. Принциповою особливістю зазначених моделей є явне врахування в них проблематики забезпечення конфіденційності та знеособленості даних. Для демонстрації можливості практичного застосування запропонованої моделі метаданих наведено фрагмент опису моделі поняття «статистичне спостереження». Планується, що розроблені моделі будуть використані під час побудови Інтегрованої системи обробки статистичних даних України.

Таблиця 5 – Фрагмент метаопису «статистичного спостереження»

опис метаданих			реквізити
рівень	код (під) групи	тип компоненти	
синт.	К1	α+β+γ+τ	назва
прагм.	К2.3 / доступність	α	відповідальний підрозділ Держкомстату
прагм.	П1	α+β+γ+τ	методика, методичні положення, роз'яснення (вказівки, рекомендації, посібники)
синт.	С1	α	код статистичного спостереження
сем.	К3, П3	γ	вид статистичного спостереження за ступенем охоплення одиниць (суцільне, несуцільне – вибіркове, основного масиву, монографічне, анкетне, моніторинг)
прагм.	К3, П3	τ	вид статистичного спостереження за часом реєстрації даних (поточне, періодичне, одноразове)
сем.	К3, П3	γ	спосіб статистичного спостереження (безпосередній облік, документальний облік, опитування – експедиційний спосіб, самореєстрація, кореспондентський спосіб)
сем.	К1	α	об'єкт статистичного спостереження
прагм.	П3	β	одиниця статистичної сукупності
прагм.	К3, П3	β	одиниця статистичного спостереження
сем.		γ	національні та локальні статистичні класифікації (назва, підстава і дата введення в дію)
синт.	С1	α	код джерела отримання даних
прагм.	С2	α+β+γ+τ	контролі повноти даних
прагм.	С2	α+β+γ+τ	арифметичні контролі даних
прагм.	С2	α+β+γ+τ	логічні контролі даних
прагм.	С2	γ	перевірки правильності кодування даних
прагм.	С2	α+β+γ+τ	алгоритм обчислення розрахункових показників та агрегація
конф.	П4.1	α+β+γ+τ	контролі дотримання правил конфіденційності даних
конф.	П4.2	α+β+γ+τ	перевірка забезпечення групової анонімності в мікрофайлах
прагм.	П2	α+β+γ+τ	оцінка якості даних статистичного спостереження

Література

1. Development of State Statistics System for Monitoring Social & Economic Transformation Project [Електронний ресурс] / The World Bank. – Режим доступу : http://web.worldbank.org/external/projects/main?Projectid=P076338&theSitePK=40941&piPK=64290415&pagePK=64283627&menuPK=64282134&Type=Overview
2. Sundgren B. Information systems architecture for national and international statistical offices. Guidelines and recommendations // Conference of European statisticians. Statistical standards and studies. – Geneva : United Nations Statistical Commission, 1999. – № 51. – 56 p.
3. SDMX Standards Version 2.0 [Електронний ресурс] – Режим доступу : http://sdmx.org
4. The OECD Data and Metadata Reporting and Presentation Handbook / Denis Ward (ed.). – Paris : OECD Publications, 2007. – 161 p.
5. Generic Statistical Business Process Model, Version 4.0 (April 2009) / Steven Vale (ed.). – METIS, 2009. – 28 p.
6. Guidelines for Statistical Metadata on the Internet // Conference of European statisticians. Statistical standards and studies. – Geneva : United Nations Statistical Commission, 2000. – № 52. – 12 p.
7. Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data, ETS № 108, 28.01.1981 (with Amendments approved by the Committee of Ministers on 15 June 1999). – Strasbourg : The Council of Europe.
8. Григорьев В.В. Обеспечение конфиденциальности статистических данных в странах Европейского Сообщества / В.В. Григорьев, Е.В. Кузнецова // Вопросы статистики. – 2000. – № 12. – С. 8-14.
9. Privacy-Preserving Data Mining: Models and Algorithms / C.C. Aggarwal, P.S. Yu (eds.). – New York : Springer, 2008. – 513 p.
10. Task Force on Metadata. Summary Report, June, 1999 [Електронний ресурс] / American Library Association. – Режим доступу : http://www.libraries.psu.edu/tas/jca/ccda/tf-meta3.html
11. Eurostat. Data Life Cycle Object Model for Eurostat Production Systems, Cronos Technologies, Unit A1, DG ESTAT, European Commission, Luxemburg, June 2002. – 25 p.
12. The Data Documentation Initiative Specification, version 3.1 (October 2009) [Електронний ресурс] – Режим доступу : http://www.ddialliance.org
13. Закон України «Про державну статистику», № 2614-XII, остання редакція від 31.03.2009 [Електронний ресурс] / Верховна Рада України. – Режим доступу : http://zakon.rada.gov.ua/cgi-bin/laws/main.cgi?nreg=2614-12
14. Чертов О.Р. Застосування вейвлет-перетворень для забезпечення групової анонімності даних / О.Р. Чертов // Системи обробки інформації. − 2010. − Вип. 3 (84). − С. 90-95.
15. Guidelines for the modeling of statistical data and metadata // Conference of European statisticians. Methodological material. – Geneva : United Nations Statistical Commission, 1995. – 30 p.

О.Р. Чертов

Моделирование метаданных описания статистической информации

В статье рассматриваются вопросы, которые возникают в ходе моделирования метаданных, используемых для описания разнообразной статистической информации. Конкретизируется известная модель статистических бизнес-процессов и предлагается новая семиотическая модель метаданных. В них, в частности, учитываются аспекты обеспечения обезличенности и конфиденциальности статистической информации, а также ее защита от раскрытия при использовании методов интеллектуального анализа данных.

Oleg Chertov

Metadata Modeling of Statistical Information Description

In the paper, we discuss the problems which arise during the modeling of the metadata used for describing various pieces of statistical information. We concretize the well-known model of statistical business processes, and propose a novel semeiotic metadata model. Both models take into account aspects of providing statistical information depersonalization and confidentiality, and also its protection against the disclosure by means of data mining methods.

Стаття надійшла до редакції 29.06.2010.